HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P.O. Box 272400
Fort Collins, Colorado 80527-2400

PATENT APPLICATION

ATTORNEY DOCKET NO. 10007456 -1

AF 2171

IN THE
UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s): Nina Mishra

Confirmation No.: 2513

Application No.: 10/039617

Examiner: Sana Al-Hashemi

Filing Date: Jan 04, 2002

Group Art Unit: 2171

Title: Computer Implemented, Fast, Approximate Clustering Based On Sampling

Mail Stop Appeal Brief-Patents
Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

TRANSMITTAL OF APPEAL BRIEF

Sir:

Transmitted herewith is the Appeal Brief in this application with respect to the Notice of Appeal filed on **April 29, 2005**.

The fee for filing this Appeal Brief is (37 CFR 1.17(c)) $500.00.

(complete (a) or (b) as applicable)

The proceedings herein are for a patent application and the provisions of 37 CFR 1.136(a) apply.

( ) (a) Applicant petitions for an extension of time under 37 CFR 1.136 (fees: 37 CFR 1.17(a)-(d) for the total number of months checked below:

( )	one month	$120.00
( )	two months	$450.00
( )	three months	$1020.00
( )	four months	$1590.00

( ) The extension fee has already been filled in this application.

(X) (b) Applicant believes that no extension of time is required. However, this conditional petition is being made to provide for the possibility that applicant has inadvertently overlooked the need for a petition and fee for extension of time.

Please charge to Deposit Account **08-2025** the sum of **$500.00**. At any time during the pendency of this application, please charge any fees required or credit any over payment to Deposit Account 08-2025 pursuant to 37 CFR 1.25. Additionally please charge any fees to Deposit Account 08-2025 under 37 CFR 1.16 through 1.21 inclusive, and any other sections in Title 37 of the Code of Federal Regulations that may regulate fees. A duplicate copy of this sheet is enclosed.

(X) I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Alexandria, VA 22313-1450. Date of Deposit: **6/27/2005**

OR

( ) I hereby certify that this paper is being transmitted to the Patent and Trademark Office facsimile number_____ on _____

Number of pages: **18**

Typed Name: **Be Henry**

Signature: Be Henry

Respectfully submitted,

Nina Mishra

By _____

Phil Lyren

Attorney/Agent for Applicant(s)
Reg. No. **40,709**

Date: **6/27/2005**

Telephone No.: **281 514 8236**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Nina Mishra, et al.	Examiner: Sana Al-Hashemi
Serial No.:	10/039,617	Group Art Unit: 2171
Filed:	January 4, 2002	Docket No.: 10007456-1
Title:	Computer Implemented, Fast, Approximate Clustering Based on Sampling	

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Appeal Brief is filed in response to the Final Office Action mailed March 1, 2005 and the Notice of Appeal filed on April 29, 2005.

AUTHORIZATION TO DEBIT ACCOUNT

It is believed that no extensions of time or fees are required, beyond those that may otherwise be provided for in documents accompanying this paper. However, in the event that additional extensions of time are necessary to allow consideration of this paper, such extensions are hereby petitioned under 37 C.F.R. § 1.136(a), and any fees required (including fees for net addition of claims) are hereby authorized to be charged to Hewlett-Packard Development Company's deposit account no. 08-2025.

I. REAL PARTY IN INTEREST

The real party-in-interest is the assignee, Hewlett-Packard Company, a Delaware corporation, having its principal place of business in Palo Alto, California.

II. RELATED APPEALS AND INTERFERENCES

There are no known related appeals or interferences known to appellant, the appellant's legal representative, or assignee that will directly affect or be directly affected by or have a bearing on the Appeal Board's decision in the pending appeal.

III. STATUS OF CLAIMS

Claims 1 – 10 stand finally rejected. No claims have been allowed. The final rejection of claims 1 – 10 is appealed.

IV. STATUS OF AMENDMENTS

Claim 1 – 10 were not amended during prosecution. Thus, the claims on appeal and in the following Claim Appendix correspond to the claims as originally filed.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The summary is set forth in three exemplary embodiments that correspond to independent claims 1, 6, and 9. Discussions about elements and recitations of these claims can be found at least at the cited locations in the specification and drawings.

Claim 1

Claim 1 recites a computer implemented method for center-based clustering a set, *S*, of *n* points to identify *k* centers through sampling of large data sets, wherein *k* is an integer value greater than one (page 5, lines 1-25; page 8, lines 14-23; FIG. 2, #200 and FIG. 4: page 10, lines 15-17; and page 20, lines 16+). The method comprising the steps of determining at least one representational value of a diameter of a space *M* that comprises said set *S* of said *n* points (FIG. 1: page 9, lines 4-10; and FIG. 2,

#230: page 10, lines 22+); obtaining a sample *R* from said set *S* of said *n* points (page 9, lines 11-21; FIG. 2, #250: page 10, lines 22+; and page 12, lines 3+); determining at least one cluster for said sample *R* (page 9, lines 11-21; FIG. 2, #280; and page 12, lines 17-19); and outputting centers, $c_1,..,c_k$, as identified by said cluster of said sample, *R* (page 10, lines 7-10; FIG. 2, #295: and page 12, line 20.).

Claim 6

Claim 6 recites a computer implemented method for assessing a quality of conjunctive clusters $t_1,..,t_k$, (page 13, lines 9-20; and page 15, lines 3-12). The method comprises the steps of: determining a length of each respective conjunction, t_i (page 15, lines 8-12); determining a probability of each respective conjunction, t_i (page 15, lines 8-12); summing the product of the length of the conjunction, t_i, with the probability of t_i for *i* ranging from 1to *k* (page 15, lines 18+); wherein, the *k* conjunctions, $t_1,..,t_k$, cover all but γ of the distribution (page 14, lines 1-16); and wherein maximizing the summing said product step optimizes conjunctive clusters (page 13, lines 14-20).

Claim 9

Claim 9 recites a computer implemented method for disjoint conjunction clustering a set *S* of *n* points through sampling of large data sets (page 13, lines 4+; and FIG. 3, #300). The method comprising the steps of: obtaining a sample, *R*, from said set, *S* (FIG. 3, #302; and page 18, line 16); generating a plurality of signatures of *k* disjoint conjunctions (page 16, lines 18+); enumerating over an each individual signature *q* of said plurality of signatures of *k* disjoint conjunctions (FIG. 3, #305: and page 18, lines 16+), by: partitioning said *R* into buckets $B_1,..,B_k$ according to said signature, *q* (FIG. 3,

#320; and page 18, lines 17+); inducing additional buckets as needed (FIG. 3, #315, #320, #325; and page 18, lines 19+); determining a conjunction t_i for each bucket of points B_i that represents the most specific conjunction that satisfies the points in B_i (FIG. 3, #330; and page 19, lines 1+); computing an empirical frequency $R(t_i)$ (FIG. 3, #340; and page 19, lines 4+); assessing a numeric quality representation from a summation of a product of the length of t_i and the empirical frequency $R(t_i)$, over all said buckets induced by said signature *s* (FIG. 3, #345; and page 19, lines 6+); and outputting *k* disjoint conjunctions of said sample *R* that exhibits a highest absolute value of said numeric quality representation (FIG. 3, #350; and page 19, lines 8+).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Claims 1– 10 are rejected under 35 U.S.C. §101 as being directed to non-statutory subject matter. In particular, the Office Actions states that the claimed subject matter is directed to a "mathematical algorithm."

VII. ARGUMENT

The rejection of claims 1– 10 under 35 USC § 101 as being directed to non-statutory subject matter is improper. Applicants respectfully requests withdraw of this rejection.

Applicants present separate arguments for each of the independent claims 1, 6, and 9.

Overview of Law: § 101

Under 35 USC § 101, patentable subject matter must have two basic criteria. First, the subject matter must be one of processes, machines, manufacturers, and compositions of matter. Generally, three categories are not included as patentable subject matter: (1) abstract ideas, (2) laws of nature (example, mathematical algorithms/equations), and (3) natural phenomena. Second, the subject matter to be patented must be "useful." Applicants' claimed subject matter meets both of these criteria.

Overview of Issues

The first issue is whether the "processes" of claims 1, 6, and 9 fall within the excluded patentable subject matter of (1) abstract ideas, (2) laws of nature, and (3) natural phenomena. The Office Action contends that the claims are directed to non-statutory subject matter of a "mathematical algorithm." Applicants respectfully disagree. Applicants respond with at least two arguments: First, the claims are not directed to a mathematical algorithm, and the claims are limited to a practical application. Second, even assuming *arguendo* that the claims do contain a mathematical algorithm, the claims do not consist solely of mathematical operations.

The second issue is whether the "processes" of claims 1, 6, and 9 produce a useful, concrete, and tangible result to have a practical application.

Issue 1: Claims within Subject Matter of § 101

The Examiner contends that the independent claims are directed to a mathematical algorithm and, therefore, are not statutory under 35 USC § 101 (see FOA at p. 2). For several reasons, Applicants respectfully disagree.

First, Applicants contend that independent claims 1, 6, and 9 recite "a computer implemented method." As clearly supported in the law, computer implemented methods are "processes" according to 35 USC § 101 (see MPEP § 2106, IV, A: "The term 'process' means process, art, or method ...."). Thus, Applicants argue claims 1 – 10 as being within the category of "processes" according to 35 USC § 101.

Second, Applicants contend that independent claims 1, 6, and 9 do not even recite mathematical equations. Applicants respectfully ask the Appeal Board to review claims 1, 6, and 9. Each of these claims recites a computer implemented method having numerous elements for clustering data. None of the elements in independent claims 1, 6, and 10 recite a mathematical equation. In other words, mathematical equations or algorithms are not even recited in the independent claims. For at least this reason, Applicants respectfully ask the Appeal Board to overrule the position of the Examiner.

Third, even assuming *arguendo* that independent claims 1, 6, and 9 recite mathematical algorithms (which they do not), these claims are still statutory under 35 USC § 101. The law clearly states that mathematical algorithms can form part of statutory subject matter. The Supreme Court stated:

> Their process admittedly employs a well-known mathematical equation, but they do not seek to pre-empt the use of that equation. Rather, they seek only to foreclose from others the use of that equation in conjunction with all of the other steps in their claimed process .... Our earlier opinions lend support to our present conclusion that a claim drawn to subject matter otherwise statutory does not become nonstatutory simply because it uses a mathematical formula, computer program, or digital computer. (*Diamond v. Diehr*, 450 U.S. 175 (1981): text omitted from quotation.)

Further, claims define non-statutory processes if they either (1) consist **solely** of mathematical operations **without some** claimed practical application, or (2) **simply** manipulate abstract ideas **without some** claimed practical application (see MPEP 2106).

Stated another way, claims define a statutory process if they are limited to a practical application within the technological arts (see MPEP 2106 or *Diamond v. Diehr*, 450 U.S. 175 at 183-184 (1981)).

In the next section (Issue 2), Applicants demonstrate that the claims neither (1) consist solely of mathematical operations without some claimed practical application nor (2) simply manipulate abstract ideas without some claimed practical application. In short, Applicants demonstrate in the next section that the independent claims produce a useful, concrete, tangible result.

Issue 2: Claims Produce Useful, Concrete, Tangible Result

The Examiner contends that the independent claims have no practical application and, thus, do not produce a useful, concrete, tangible result. Instead, the Examiner argues that the claims are to a mathematical algorithm and, therefore, are not statutory under 35 USC § 101 (see FOA at p. 2). For several reasons, Applicants respectfully disagree.

Applicants' claims have a practical application in the technological arts since the claims produce a concrete, tangible, and useful result. In other words, the claims recite at least one step or one act that produces something that is concrete, tangible, and useful. Applicants provide examples for each of the independent claims 1, 6, and 9.

Claim 1 Produces Useful, Concrete, Tangible Result

Claim 1 recites a computer implemented method for center-based clustering to identify *k* centers. The last claim element recites: "outputting centers, $c_1,\ldots,c_k$, as identified by said cluster of said sample, R." Thus, claim 1 **outputs** centers that are identified by said cluster of said sample. In other words, the claim recites a concrete, tangible, and useful result as an output of centers.

Claim 1 thus provides a "real world" value (i.e., the output centers) for a computer implemented method. This real world value is more than a mere idea or concept. Further, the output of claim 1 proves that the claimed process does not consist **solely** of the manipulation of an abstract idea. By contrast, the claim provides a concrete and tangible result.

Claim 6 Produces Useful, Concrete, Tangible Result

Claim 6 recites a computer implemented method for assessing a quality of conjunctive clusters. The claim recites (portions omitted):

> summing the product of the length of the conjunction, t_i, with the probability of t_i for i ranging from 1 to k;
>
> wherein maximizing the summing said product step optimizes conjunctive clusters.

In order to assess a quality of clusters, claim 6 recites several elements. In one element, a product of length of the conjunction *t* is summed with a probability t_i. Then, this result is maximized to optimize the conjunctive clusters. Thus, claim 6 optimizes conjunctive clusters. In other words, the claim recites a concrete, tangible, and useful result as optimizing conjunctive clusters to assess the quality of the conjunctive clusters.

Claim 6 thus provides a "real world" value (i.e., optimization of conjunctive clusters) for a computer implemented method. This real world value is more than a mere idea or concept. Further, the optimizing result of claim 6 proves that the claimed process does not consist **solely** of the manipulation of an abstract idea. By contrast, the claim provides a concrete and tangible result.

Claim 9 Produces Useful, Concrete, Tangible Result

Claim 9 recites a computer implemented method for clustering through sampling a large data set. The last claim element recites:

> outputting k disjoint conjunctions of said sample R that exhibits a highest absolute value of said numeric quality representation.

Thus, claim 9 **outputs** disjoint conjunctions that exhibit a highest absolute value of a numeric quality representation. In other words, the claim recites a concrete, tangible, and useful result as an output of disjoint conjunctions.

Claim 9 thus provides a "real world" value (i.e., the output of disjoint conjunctions) for a computer implemented method. This real world value is more than a mere idea or concept. Further, the output of claim 9 proves that the claimed process does not consist **solely** of the manipulation of an abstract idea. By contrast, the claim provides a concrete and tangible result.

Law Supports Position of Applicants

The legal position of the Applicants is clearly supported in MPEP 2106 and case law, such as *AT&T Corp. v. Excel Communications,*, 172 F.3d 1352 (Fed. Cir. 1999). For example, the MPEP clearly states: "Only when the claim is **devoid** of any limitation to a practical application in the technological arts should it be rejected under 35 USC 101" (MPEP 2106: Emphasis added). Applicants have shown that the independent claims are not devoid of any limitation to a practical application in the technological arts. As noted, claim 1 recites a real world value (i.e., the output centers); claim 6 recites a real world value (i.e., optimization of conjunctive clusters); and claim 9 recites a real world value (i.e., the output of disjoint conjunctions).

Next, Applicants respectfully cite MPEP 2106 to support further their position:

> The applicant is in the best position to explain why an invention is believed useful. Office personnel should therefore focus their efforts on pointing out statements made in the specification that identify all practical applications for the invention. Office personnel should **rely** on such statements throughout the examination when assessing the invention for compliance with all statutory criteria. An applicant may assert more than one practical application, but **only one is necessary to satisfy the utility requirement**. Office personnel should review the entire disclosure to determine the features necessary to accomplish at least one asserted practical application. (Bold added).

If the Appeal Board doubts the usefulness of the claimed invention, then Applicants respectfully ask the Appeal Board to read "Background of the Invention." Here, Applicants discuss the numerous problems associated with the field of computer implemented clustering techniques. The background describes one exemplary problem:

> Essentially, most prior art clustering methods are not designed to work with massively large datasets, especially because most computer implemented clustering methods require multiple passes through the entire datasets which may overwhelm or bog down a computer system if the dataset is too large. As such, it may not be feasible to cluster large datasets, even given the recent developments in large computing power. (Page 1, line 24 – page 2, line2).

Applicants' specification discusses numerous other examples of practical applications for the claimed invention. As another example, the specification clearly discusses a practical application to data mining and computer implemented clustering techniques:

> The application of clustering to knowledge discovery and data mining require a clustering technique with quality and performance guarantees that apply to large datasets.... As described fully below, the fast sampling technique of the present invention is sublinear, and as such, significantly improves the efficiency of computer resources, reduces time of execution, and ultimately provides for an accurate, fast technique for clustering which is independent of the size of the data set. (Page 5, lines 11-22: text omitted).

Applicants submit that the specification discusses numerous other examples of practical applications. As one example, Applicants respectfully ask the Appeal Board to read the section entitled "Computer Implementation Efficiency" beginning on page 19.

Applicants cite the noted exemplary portions of the Background and Specification to show that when the claims are given their broadest reasonable interpretation in light of the specification and those skilled in the art, the claims clearly recite a useful, tangible, concrete result. In other words, claim 1 recites a practical application with a useful result (i.e., the output centers); claim 6 recites a practical application with a useful result (i.e., optimization of conjunctive clusters); and claim 9 recites a practical application with a useful result (i.e., the output of disjoint conjunctions).

No Preemption of Ideas with Claims

Under 35 USC 101, laws of nature (such as mathematical formula) are not patentable. One rationale is courts have expressed concern over "preemption" of laws of

nature (see MPEP § 2106, IV, A.). "In fact, such concerns are only relevant to claiming a scientific truth or principle" (Id.). In other words, the courts do not want patent holders to have exclusive rights to laws of nature or abstract ideas. Applicants contend that these concerns are not warranted with the claims of the present invention. In other words, the claims do not seek an exclusive right to a law of nature (i.e., a mathematical equation) or abstract idea. Instead, each independent claim recites elements that do not form a mathematical equation, and each independent claim produces a useful, concrete, tangible result with a practical application. Appellants respectfully ask the Appeal Board to review elements as a whole of each independent claim 1, 6, and 9 (see *Diamond v. Diehr*, 450 U.S. 175 (1981) at 188: "In determining the eligibility of respondent's claimed process for patent protection under § 101, their claims must be considered as a whole."). Independent claims 1, 6, and 9 do not even recite laws of nature (e.g., mathematical equations, such as $E = mc^2$), abstract ideas, or natural phenomenon.

Burdon on Examiner for Prima Facie Case

Applicants respectfully argue that the Examiner has the burden to establish that independent claims 1, 6, and 9 do not meet the statutory requirements of 35 USC § 101. In fact, the MPEP is very clear on this burden:

> Office personnel have the burden to establish a *prima facie* case that the claimed invention as a whole is directed to solely an abstract idea or to manipulation of abstract ideas or does not produce a useful result. Only when the claim is devoid of any limitation to a practical application in the technological arts should it be rejected under 35 U.S.C. 101. Compare *Musgrave*, 431 F.2d at 893, 167 USPQ at 289; *In re Foster*, 438 F.2d 1011, 1013, 169 USPQ 99, 101 (CCPA 1971). Further, when such a rejection is made, Office personnel must expressly state how the language of the claims has been interpreted to support the rejection. (MPEP § 2106, II, A).

Applicants respectfully submit that the Examiner has not established this *prima facie* case.

CONCLUSION

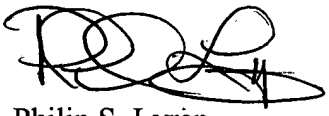
For at least the aforementioned reasons, independent claims 1, 6, and 9 are directed to statutory subject matter under 35 USC § 101. A dependent claim inherits the limitations of a base. Thus, for at least the reasons given in connection with independent claims 1, 6, and 9, the respective dependent claims are also directed to statutory subject matter under 35 USC § 101.

In view of the above, Applicants respectfully request the Board of Appeals to reverse the Examiner's rejection of all pending claims.

Any inquiry regarding this Amendment and Response should be directed to Philip S. Lyren at Telephone No. (281) 514-8236, Facsimile No. (281) 514-8332. In addition, all correspondence should continue to be directed to the following address:

Hewlett-Packard Company
Intellectual Property Administration
P.O. Box 272400
Fort Collins, Colorado 80527-2400

Respectfully submitted,



Philip S. Lyren
Reg. No. 40,709
Ph: 281-514-8236

CERTIFICATE UNDER 37 C.F.R. 1.8: The undersigned hereby certifies that this paper or papers, as described herein, are being deposited in the United States Postal Service, as first class mail, in an envelope address to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on this 27th day of June, 2005.

By Be Henry
Name: Be Henry

VIII. Claims Appendix

1. A computer implemented method for center-based clustering a set, *S*, of *n* points to identify *k* centers through sampling of large data sets, wherein *k* is an integer value greater than one, the method comprising the steps of:

determining at least one representational value of a diameter of a space *M* that comprises said set *S* of said *n* points;

obtaining a sample *R* from said set *S* of said *n* points;

determining at least one cluster for said sample *R*; and

outputting centers, $c_1,..,c_k$, as identified by said cluster of said sample, *R*.

2. The method as set forth in claim 1, further comprising the step of reducing the number of dimensions *d* to log *n*, if *d* is larger than log *n*, prior to determining said representational value of said diameter of said space *M*.

3. The method as set forth in claim 2, further comprising the steps of:

executing a discrete clustering of sample *R* in a reduced space and

translating said centers back to original space prior to outputting them.

4. The method as set forth in claim 1, wherein the size of the sample *R* is greater than or equal to the resulting value of:

$$O((\frac{M\alpha}{\varepsilon})^2(dk\ln\frac{12dM}{\varepsilon}+\ln\frac{4}{\delta}))$$ if R is in Euclidean space and

$$O((\frac{M\alpha}{\varepsilon})^2(k\ln n+\ln\frac{4}{\delta})$$ if R is in a metric space.

5. The method as set forth in claim 1, wherein the step of determining said diameter *M*, if *M* is unknown, comprises the step of obtaining a sample of size greater than or equal to the resulting value of:

$$\frac{2d}{\varepsilon}\log\frac{2d}{\delta}.$$

6. A computer implemented method for assessing a quality of conjunctive clusters $t_1,..,t_k$, comprising the steps of:

determining a length of each respective conjunction, t_i;

determining a probability of each respective conjunction, t_i;

summing the product of the length of the conjunction, t_i, with the probability of t_i for *i* ranging from 1to *k*;

wherein, the *k* conjunctions, $t_1,..,t_k$, cover all but γ of the distribution; and

wherein maximizing the summing said product step optimizes conjunctive clusters.

7. The method set forth in claim 6, wherein the length of each respective said conjunction is determined by determining a number of variables in each respective said conjunction.

8. The method set forth in claim 7, wherein the step of determining a probability of each respective conjunction, t_i, comprises the step of determining a number of points that satisfy said conjunctions.

9. A computer implemented method for disjoint conjunction clustering a set *S* of *n* points through sampling of large data sets, the method comprising the steps of:

obtaining a sample, *R*, from said set, *S*;

generating a plurality of signatures of *k* disjoint conjunctions;

enumerating over an each individual signature *q* of said plurality of signatures of *k* disjoint conjunctions, by:

partitioning said *R* into buckets $B_1,..,B_k$ according to said signature, *q*;

inducing additional buckets as needed;

determining a conjunction t_i for each bucket of points B_i that represents the most specific conjunction that satisfies the points in B_i;

computing an empirical frequency $R(t_i)$;

assessing a numeric quality representation from a summation of a product of the length of t_i and the empirical frequency $R(t_i)$, over all said buckets induced by said signature *s*; and

outputting *k* disjoint conjunctions of said sample *R* that exhibits a highest absolute value of said numeric quality representation.

10. The method as set forth in claim 9 wherein the size of the sample R is greater than or equal to:

$$\min\{\frac{1}{\gamma}(\text{dkln } 3 + \ln\frac{2}{\delta}), \frac{2\text{d2k2}}{\varepsilon 2}(\text{dln } 3 + \ln\frac{2}{\delta})\}.$$

IX. EVIDENCE APPENDIX

None.

X. RELATED PROCEEDINGS APPENDIX

None.